DOCUMENT RESUME

ED 052 295 VT 010 867

TITLE	Study to Compare Reliability of Performance on Live and Recorded Dictation Tests.
INSTITUTION	Manpower Administration (DOL), Washington, D.C. U.S. Training and Employment Service.
REPORT NO	USTES-TR-28a
PUB DATE	Dec 69
NOTE	42p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Comparative Analysis, Comparative Testing, *Evaluation, *Performance Tests, *Stenography, *Testing, *Test Reliability
IDENTIFIERS	*Utah

ABSTRACT

To compare the reliability of performance on recorded dictation tests with performance on live tests, 216 university students who were nearing completion of an intermediate shorthand course and 26 job applicants seeking stenographic positions were divided into 10 groups, with five receiving live dictation and five receiving recorded dictation. The experimental test battery consisted of three 80-word-per-minute dictation exercises, administered in 50-minute versions by trained examiners during May and June, 1966. Some conclusions were: (1) Either live or recorded dictation test administration is acceptable to examinees, (2) Most examinees prefer taking dictation from a male voice, (3) An analysis of student sample variance verifies that experience rather than test form or method of instruction exerted the greatest influence on shorthand skill, (4) As measured by Pearson product-moment correlations, there is no consistent relationship between test performance and the personal factors of age, education, amount of shorthand training, and length of shorthand experience, and (5) Although students had less transcription time than job applicants, their mean error scores were considerably lower. (SB)

58

December 1969 United States Training and Employment Service Technical Report 28a

ED052295

Study to Compare Reliability of Performance on Live and Recorded Dictation Tests

VT010867

U.S. DEPARTMENT OF LABOR
MANPOWER ADMINISTRATION

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

1

ED052295

USTES Test Research Report No. 28a

Study to Compare Reliability of Performance on Live and Recorded Dictation Tests

Manpower Administration

U.S. Department of Labor

December 1969

FOREWORD

Extensive research conducted under the Federal-State cooperative test research program in the Training and Employment Service has led to the development of many tools useful in vocational counseling and placement. These tools include aptitude tests, proficiency tests, and noncognitive measures based on instruments such as interest inventories and biographical information blanks.

The purpose of this series of reports is to provide results of significant test research projects as they are completed. These reports will be of interest to users of the tests and to test research personnel in other organizations.

This report was prepared in the Utah Department of Employment Security as part of a cooperative research study under the general direction of the U.S. Training and Employment Service. It is one of seven prepared by State agencies affiliated with the U.S. Training and Employment Service on the subject of live and recorded methods of dictation test administration. The seven studies conducted from 1959 to 1967 are summarized in the USTES Test Research Report No. 28, December 1969.

Study to Compare Reliability of Performance
on Live and Recorded Dictation Tests

Contents

	Page
Purpose of Study	1
Sample Description	1
Experimental Test Battery	6
Data Collection	6
Data Processing	9
Data Analysis	10
Sample Characteristics	10
Test Administration Preferences	15
Test Performance	18
Test Performance and Personnel Factors	25
Reliability of Testing Methods	28
Summary and Conclusions	34

STUDY TO COMPARE RELIABILITY OF PERFORMANCE ON LIVE AND RECORDED DICTATION TESTS

PURPOSE OF STUDY

This study was undertaken to compare the reliability of performance on recorded dictation tests and the reliability of performance on these same tests when administered "live."

SAMPLE DESCRIPTION

The total research sample was drawn from two sources. These were: (1) University shorthand students and (2) local office job applicants seeking stenographic positions. Preliminary plans to include employed stenographers and business school students enrolled in shorthand courses were eliminated as the study progressed. Sample members were administered two equivalent test forms either "live" or by tape recording.

The University sample was composed of 227 students who were nearing completion of Business Education Course 112, Intermediate Shorthand at Brigham Young University (BYU), Provo, Utah. Content of this semester course is described in the University Catalog as follows: "Development of shorthand writing and control beginning with 60 words per minute to develop a rate of 80 to 100 words per minute for three minutes on unfamiliar material." Business Education 111, Elementary Shorthand is a prerequisite. Female instructors teach the Business Education 112 Course, but recorded dictation exercises with mostly male voices are used extensively in the teaching process.

Three classes of Business Education Course 112 students were tested. Each of the three classes was divided into several sections. Each section had its

own instructor and regularly scheduled classroom and instruction period. Classroom facilities for the various sections were comparable. The experimental test battery was administered on May 11 and May 12, 1966 to students in their own sections during their regularly scheduled class periods of 8:00 a.m., 1:00 p.m. or 2:00 p.m. The method of administration -- "live" or "recorded" -- for each section was selected randomly. Four trained test examiners from the Utah Department of Employment Security administered the dictation tests. Two of these examiners were males and two were females.

Nine students were dropped from the sample because they did not attempt to transcribe all of the dictation exercises. Two more students were eliminated because of incomplete personnel information. Therefore, the final University sample consisted of 216 students all of whom were females except one. This sample was about equally divided between the two methods of administration. There were 105 students in the "live" group and 111 in the "recorded" group.

The job applicant sample was obtained from the Professional, Clerical and Sales Employment Office (PCS) in Salt Lake City, Utah. Persons seeking stenographic positions were administered the experimental test battery in the employment office proficiency testing room during regularly scheduled testing sessions between June 6 and June 23, 1966. The method of administration was changed randomly from day to day, but the method selected for a given day remained the same for all testing sessions. Two trained female test examiners from the Utah Department of Employment Security administered the dictation tests to 28 job applicants. Two persons were eliminated from this group. One applicant did not complete transcription, and the other was identified as a member of the BYU student sample. The final sample of 26 applicants included one male with 14 examinees transcribing "live" dictation tests and 12 transcribing "recorded" dictation tests.

The final research sample of 242 examinees included ten sub-samples. Sampling characteristics of these sub-samples are summarized in Table 1 below:

TABLE 1

Summary of Sampling Characteristics for the Ten Sub-Samples

Sample	N	Source	Method	Examiner Sequence*	Testing Schedule
01	25	BYU	Live	A-4, B-10	5-11-66 1:00 p.m.
02	35	BYU	Live	C-4,10	5-11-66 1:00 p.m.
03	29	BYU	Recorded	B-4,10	5-11-66 2:00 p.m.
04	33	BYU	Recorded	A-4,10	5-11-66 2:00 p.m.
05	18	BYU	Recorded	C-4,10	5-11-66 2:00 p.m.
06	34	BYU	Live	B-4, A-10	5-12-66 8:00 a.m.
07	11	BYU	Live	C-4,10	5-12-66 8:00 a.m.
08	31	BYU	Recorded	D-4,10	5-12-66 8:00 a.m.
09	14	PCS	Live	C-4,10 E-4,10 C-4,E-10 E-4,C-10	6-6-66 to 6-23-66
10	12	PCS	Recorded	C-4,10 E-4,10 C-4, E-10 E-4, C-10	6-6-66 to 6-23-66

* A, C, E = Female Examiners

B and D = Male Examiners

4 = USES Dictation Exercise No. 4

10 = USES Dictation Exercise No. 10

Table 2 shows age, education, training, and experience characteristics for the combined student sample and the combined job applicant sample. The average member of the student sample was relatively young and was nearing the end of two years in college. Age range was from 18.0 to 29.2 years with a mean of 19.9 years. Educational range was from 12 to 17 years with a mean of 13.6 years. All students had had at least one course in shorthand with an upper range of four courses.

TABLE 2

Means, Standard Deviations (SD), and Ranges for Age, Education, Training and Experience for the Combined Samples

Combined Student Sample
N=216

	Mean	SD	Range
Age (years)	19.927	1.849	18.0-29.2
Education (years)	13.611	.926	12-17
Training (courses)	2.273	.813	1-4
Experience (months)	5.689	9.824	0-48

Combined Job Applicant Sample
N=26

	Mean	SD	Range
Age (years)	20.558	5.801	17.6-42.6
Education (years)	12.692	.867	12-15
Training (courses)	2.115	.640	1-4
Experience (months)	24.692	66.117	0-276

The sample mean for shorthand training was 2.3 courses. Some students had had no on-job shorthand experience while others had worked up to 48 months on jobs where shorthand was used. Average shorthand work experience was 5.7 months. Except for shorthand experience, the variability of each of these student characteristics was limited as evidenced by standard deviation values.

The combined job applicant sample of 26, although much smaller, was not too different from the combined student sample in mean age, education, and number of training courses. On the average, the job applicant sample was .7 years older, had .9 years less education, and had completed .2 fewer shorthand courses. However, the average job applicant had had more than four times as much on-job shorthand experience. The job applicant sample also showed more variability in age and experience than the student sample.

For purposes of analysis the eight BYU sub-samples were combined along specific research dimensions. These dimensions included number of examiners, method of administration, some stenographic experience, and no stenographic experience. These configurations are shown below:

Live		Recorded	
One Examiner	(N = 46)	One Examiner	(N = 111)
Two Examiners	(N = 59)	Two Examiners	(N = 0)
Some Experience	(N = 50)	Some Experience	(N = 42)
No Experience	(N = 55)	No Experience	(N = 69)
Total	(N = 105)	Total	(N = 111)

EXPERIMENTAL TEST BATTERY

Three USES 80 words per minute dictation exercises were administered either "live" or by tape recording to all members of the research sample. These included a Practice Exercise and Exercise No. 4 and Exercise No. 10. All of these exercises were taken from the Guide to the Use of Typing, Dictation and Spelling Tests (hereafter referred to as the Guide), published by the Bureau of Employment Security in December, 1953.

Exercise sequence was always the same for both "live" and "recorded" test administration. This sequence was: (1) Administration of Practice Exercise; (2) Administration of Exercise No. 4; (3) Transcription of Exercise No. 4; (4) Administration of Exercise No. 10; and (5) Transcription of Exercise No. 10. The exercises were read verbatim from the Guide by each test examiner when administering "live" dictation tests. "Recorded" dictation tests utilized standard tape recordings of these same exercises which were recorded professionally by sound engineers using the male voice of an experienced radio broadcaster. These standard recordings were played by each examiner when administering "recorded" dictation tests.

All test instructions were given "live." Standard directions were used when administering the dictation exercises, but some information was added to these standard instructions to facilitate administration by tape recording. An introductory statement explaining the purpose of the study preceded the giving of test instructions to all examinees.

DATA COLLECTION

Four trained examiners, two male and two female, administered the experimental battery to the BYU student sample. Some students were administered the exercises "live" and some by standard tape recording. Correct timing of "live"

exercises was maintained by using a stop watch. Only one examiner was present in a student section when the exercises were administered by tape recording. Taped exercises were played from a central location and piped into appropriate classrooms upon signal from the individual examiner. With this arrangement the same recorded exercise could be administered simultaneously to more than one section whenever the test schedule so required. Either one or two examiners were present in each student section when the exercises were administered "live." When two examiners assisted in administration, one examiner gave Exercise No. 4; and the other examiner gave Exercise No. 10. The two examiners then exchanged exercises when administering dictation tests to a different section.

Because test administration had to be limited to a 50-minute class period, no break was allowed in the testing sequence; and the standard transcription time for Exercises4 and 10 was reduced from 20 minutes to 15 minutes each. Transcriptions were typed double space by students while seated at individual typing stations. Each examinee recorded his name, school, test date, dictation speed, method of administration, and name of examiner(s) in the upper right hand corner of each transcription sheet. Completed transcriptions were collected and maintained in batches by student section.

Prior to the actual testing period each student was asked to complete an information sheet. This sheet provided space for entering personnel information such as name, date of birth, highest year of education completed, length of shorthand training, and length of shorthand work experience. It also asked the student's personal preferences on method of test administration and sex of examiner. All personal and preference data were collected for the final sample of 216 students.

Two female examiners administered the tests to the job applicant sample in small groups of one to four examinees. One of these examiners was the testing clerk assigned to the PCS Office, and the other examiner was one of those who had helped administer dictation exercises to the BYU student sample. The two examiners alternated in giving the test batteries. Sometimes one examiner gave the entire battery and sometimes the other. On other occasions both examiners participated, one examiner giving Exercise No. 4 and the other Exercise No. 10. Examiner order was reversed from group to group when both examiners participated in test administration. Recorded exercises were played monaurally on a Sony Sterecorder "200" equipped with two speakers spaced about 4 feet apart. Tone and volume were preset to provide optimum quality of sound reproduction. "Live" dictation exercises were administered by having the examiner(s) read the test verbatim from the Guide, maintaining correct dictation speed by means of a stop watch.

Transcriptions of shorthand notes were typed double space by applicants while seated at individual typing stations. Twenty minutes was allowed for each transcription. Each examinee recorded his name, address, test date, dictation speed, and method of administration on his transcription sheet. Examiner identity was added by the administrator after the transcriptions were completed.

Information sheets like those used with the student sample were completed after testing by each applicant to obtain personnel data and testing preferences. Data on age, education, shorthand training, and shorthand work experience were collected for all 26 members of the final job applicant sample. Preferences for method of test administration and sex of examiner were obtained from all but one job applicant, an examinee who was administered the test "live."

DATA PROCESSING

Data for each of the ten sub-samples were readied for processing by arranging Information Sheets and transcription materials alphabetically by examinee surname. After this procedure each dictation exercise was scored twice according to "Directions for Scoring Transcription" appearing in the Guide, and the error score was posted on the face of the transcription sheet. Transcription materials for each examinee were then matched with his Information Sheet.

Testing preferences for each examinee were read from Information Sheets and posted to work sheets. Percentages of sample response were then computed for different sample combinations from tallies of preference categories. Personnel data and error scores were posted to other work tables for key punching into data cards. Postings were checked for accuracy, and all data punched into cards were verified.

Punch cards containing examinee name, sex, examinee number, study number, sample number, error scores on Exercises 4 and 10, age, education, shorthand training, and shorthand work experience were then used for ADP computation of basic statistics for these variables. These statistics included means, standard deviations, ranges, Pearson product-moment correlations, and reliability coefficients. Machine processing of data was accomplished for individual and combined samples from appropriate decks of data cards. Hard copies were printed for each machine run, and pertinent tables were prepared from these ADP printouts.

Analysis of variance and tests for significance of difference between means and reliability coefficients were computed by means of a desk calculator. All of these hand calculations were checked for accuracy.

DATA ANALYSIS

Sample Characteristics

Tables 3,4,5, and 6 show number of examinees, means, standard deviations, and ranges for age, education, shorthand training, and shorthand work experience for the ten experimental samples. Analysis of these data indicates that members of the student samples were, on the average, quite similar as far as age and education are concerned. They were generally similar in average amount of training received but varied considerably in average length of work experience. With the exception of work experience the average members of the two job applicant samples were very much alike. However, Sample 9 was a somewhat older group and exhibited greater variability in age and shorthand work experience.

Although variability in age of one of the job applicant samples (Sample 9) was greater than that of the student samples, the mean ages of the two job applicant samples were about like those of the student samples. On the average, job applicant samples had less education than student samples. Mean length of training for members of job applicant samples was more than mean length of training for some student samples but less than mean length of training for other student samples. Mean work experience of job applicant samples appeared to be higher than mean work experience of all student samples except Sample 4.

Although group differences still existed, mean differences among individual student samples tended to be less pronounced when members of the eight experimental samples were combined according to type of test administration and factor of work experience, but no real experience difference was apparent between experienced students taking the dictation tests "live" or those taking them "recorded." As measured by standard deviations, the greatest variability in age, education, training, and work experience fluctuated from sample to sample for each of these characteristics.

TABLE 3

Means, Standard Deviations (SD), and Ranges for Age (years) for the Ten Experimental Samples

Sample	N	Mean	SD	Range
1	25	20.5	2.55	18.6 - 29.2
2	35	20.2	1.55	18.0 - 24.2
3	29	19.8	1.44	18.1 - 26.0
4	33	19.7	1.08	18.2 - 23.3
5	18	19.8	1.16	18.5 - 22.9
6	34	19.5	1.10	18.1 - 23.2
7	11	20.4	1.23	18.5 - 22.0
8	31	20.1	1.59	18.4 - 24.3
9	14	21.7	7.62	17.6 - 42.6
10	12	19.2	1.39	17.8 - 22.8

TABLE 4

Means, Standard Deviations (SD), and Ranges for Education (years) for the Ten Experimental Samples

Sample	N	Mean	SD	Range
1	25	13.680	1.085	12 - 16
2	35	13.829	1.028	13 - 17
3	29	13.586	1.034	12 - 17
4	33	13.424	.740	13 - 16
5	18	13.500	.833	13 - 16
6	34	13.382	.687	13 - 16
7	11	14.273	1.135	13 - 16
8	31	13.613	.748	13 - 15
9	14	12.571	.623	12 - 14
10	12	12.833	1.067	12 - 15

TABLE 5

Means, Standard Deviations (SD), and Ranges for Shorthand Training (number of courses) for the Ten Experimental Samples

Sample	N	Mean	SD	Range
1	25	1.680	.676	1 - 3
2	35	2.457	.769	1 - 4
3	29	1.759	.727	1 - 3
4	33	2.576	.605	1 - 4
5	18	2.167	.764	1 - 3
6	34	2.824	.706	1 - 4
7	11	1.727	.862	1 - 4
8	31	2.355	.650	1 - 3
9	14	2.071	.703	1 - 4
10	12	2.167	.553	1 - 3

TABLE 6

Means, Standard Deviations (SD), and Ranges for Shorthand Work Experience (months) for the Ten Experimental Samples

Sample	N	Mean	SD	Range
1	25	3.920	8.859	0 - 38
2	35	8.000	12.495	0 - 48
3	29	.724	2.318	0 - 12
4	33	9.273	11.349	0 - 36
5	18	7.833	13.061	0 - 48
6	34	6.765	8.974	0 - 24
7	11	5.091	7.051	0 - 24
8	31	3.129	5.458	0 - 18
9	14	38.214	90.026	0 - 276
10	12	8.917	17.221	0 - 60

Tables 7,8,9, and 10 show these data for the combined student samples.

TABLE 7

Means, Standard Deviations (SD), and Ranges for Age (years) for Combined Samples of Brigham Young University Students

Sample	N	Mean	SD	Range
One Examiner	46	20.2	1.49	18.0 - 24.2
Two Examiners	59	19.9	1.92	18.1 - 29.2
All Live	105	20.1	1.75	18.0 - 29.2
All Recorded	111	19.9	1.36	18.1 - 26.0
Live-Experience	50	20.0	1.38	18.5 - 23.3
Live-No Experience	55	20.1	2.03	18.0 - 29.2
Recorded-Experience	42	20.1	1.41	18.4 - 23.6
Recorded-No Experience	69	19.7	1.31	18.1 - 26.0

TABLE 8

Means, Standard Deviations (SD), and Ranges for Education (years) for Combined Samples of Brigham Young University Students

Sample	N	Mean	SD	Range
One Examiner	46	13.935	1.071	13 - 17
Two Examiners	59	13.508	.890	12 - 16
All Live	105	13.695	.996	12 - 17
All Recorded	111	13.532	.847	12 - 17
Live-Experience	50	13.680	1.028	12 - 16
Live-No Experience	55	13.709	.966	13 - 17
Recorded-Experience	42	13.619	.872	13 - 16
Recorded-No Experience	69	13.478	.827	12 - 17

TABLE 9

Means, Standard Deviations (SD), and Ranges for Shorthand Training (number of courses) for Combined Samples of Brigham Young University Students

Sample	N	Mean	SD	Range
One Examiner	46	2.283	.851	1 - 4
Two Examiners	59	2.339	.895	1 - 4
All Live	105	2.314	.876	1 - 4
All Recorded	111	2.234	.747	1 - 4
Live-Experience	50	2.520	.854	1 - 4
Live-No Experience	55	2.127	.854	1 - 4
Recorded-Experience	42	2.500	.588	1 - 3
Recorded-No Experience	69	2.072	.786	1 - 4

TABLE 10

Means, Standard Deviations (SD), and Ranges for Shorthand Work Experience (months) for Combined Samples of Brigham Young University Students

Sample	N	Mean	SD	Range
One Examiner	46	7.304	11.498	0 - 48
Two Examiners	59	5.559	9.035	0 - 38
All Live	105	6.324	10.225	0 - 48
All Recorded	111	5.090	9.389	0 - 48
Live-Experience	50	13.280	11.277	1 - 48
Live-No Experience	55	----	----	----
Recorded-Experience	42	13.452	10.977	1 - 48
Recorded-No Experience	69	----	----	----

Test Administration Preferences

Information showing preferences for taking dictation tests was obtained from all sample members except one job applicant in Sample 9. These preferences indicated whether the examinee preferred taking dictation by "recorded" or "live" administration and from a male or female voice. "No preference" was indicated when neither alternative was dominant.

Some preference fluctuations between "recorded" and "live" administration occurred among the ten samples (see Table 11). The "no preference" alternative was selected by the greatest percentage of Samples 1, 2, 5, 6 and 10. Of the two remaining choices, the percentages preferring "live" or "recorded" dictation were about equal for Samples 1 and 6, but "recorded" dictation was the secondary choice of Samples 2 and 10. "Live" dictation was the secondary choice of Sample 5. In Sample 7, predominant preference percentages were about equally divided between the "recorded" and "no preference" categories. In Samples 3, 4, 8 and 9 the "no preference" alternative was not the primary choice. "Recorded" dictation tests were administered to Samples 3, 4 and 8. The greatest percentage of Sample 3 preferred "live" dictation, but the greatest percentage of Sample 4 and the greatest percentage of Sample 8 preferred "recorded" dictation. "Live" dictation tests were administered to Sample 9, and the largest proportion of this sample preferred this method of administration.

A grouping of subsamples presented much the same picture. Largest percentages were under the "no preference" category with no clear cut preference for one particular method of administration. Of the combined sample of 241 examinees, 29% preferred "live" administration; 30% preferred "recorded" administration and 41% indicated "no preference." Table 12 summarizes these combined sample preferences for method of dictation test administration.

TABLE 11

Sample Preferences for Method of Dictation Test Administration

Sample	N	Method	Preference Percentages		
			Live	Recorded	Either
1	25	Live	28	24	48
2	35	Live	23	34	43
3	29	Recorded	52	14	34
4	33	Recorded	21	42	37
5	18	Recorded	22	6	72
6	34	Live	32	30	38
7	11	Live	10	45	45
8	31	Recorded	26	55	19
9	13	Live	62	0	38
10	12	Recorded	8	25	67

TABLE 12

Combined Sample Preferences for Method of Dictation Test Administration

Sample	N	Method	Preference Percentages		
			Live	Recorded	Either
Student (Samples 1,2,6,7)	105	Live	26	31	43
Student (Samples 3,4,5,8)	111	Recorded	31	32	37
Combined Student (Samples 1-8)	216	Live-Recorded	28	32	40
Combined Applicant (Samples 9,10)	25	Live-Recorded	36	12	52
Combined Total (Samples 1-10)	241	Live-Recorded	29	30	41

Most examinees preferred taking dictation from a male voice. The only exceptions were Sample 9 which had no job applicant preferring a male voice and Sample 10 which had only 3 or 25% of the 12 job applicants preferring a male voice. In these two samples the majority had no voice preference. Table 13 shows the voice preference of each individual sample.

TABLE 13

Sample Preferences for Administration of Dictation Exercises by Male or Female Voice

Sample	N	Voice	Preference Percentages		
			Male	Female	Either
1	25	Male-Female	80	0	20
2	35	Female	74	9	17
3	29	Male	79	4	17
4	33	Male	76	3	21
5	18	Male	78	0	22
6	34	Male-Female	62	3	35
7	11	Female	73	0	27
8	31	Male	91	3	6
9	13	Female	0	15	85
10	12	Male	25	17	58

Grouping samples according to similarities of administration did not change these preference trends. Of the 216 students, 165 or 76% preferred a male voice; 44 or 21% had no preference, and only 7 or 3% preferred a female voice. Moreover, most of the combined job applicant sample (72%) did not have a voice preference. For all ten samples 70% preferred a male voice, 5% preferred a female voice, and 25% had no preference. These grouped data are presented in Table 14.

TABLE 14

Combined Sample Preferences for Administration of Dictation Exercises by Male or Female Voice

Sample	N	Voice	Preference Percentages		
			Male	Female	Either
Student-One Examiner (Samples 2,7)	46	Female	74	6	20
Student-Two Examiners (Samples 1,6)	59	Male-Female	69	2	29
Student-Live (Samples 1,2,6,7)	105	Male-Female	71	4	25
Student-Recorded (Samples 3,4,5,8)	111	Male	81	3	16
Combined Student (Samples 1-8)	216	Male-Female	76	3	21
Combined Applicant (Samples 9,10)	25	Female	12	16	72
Combined Total (Samples 1-10)	241	Male-Female	70	5	25

Test Performance

The various subsamples performed differently on the dictation exercises. On Exercise No. 4 rounded mean error scores ranged from 5 to 6 for Samples 4,6 and 7; 8 to 9 for Samples 2 and 5; 12 for Sample 8; 17 to 18 for Samples 9 and 10; and 20 to 24 for Samples 1 and 3. A somewhat similar pattern was shown on Exercise No. 10 except for slight improvement by Samples 5 and 7; a small increase in mean errors for Samples 4 and 6; and a noticeable increase in errors for Samples 1 and 10, making a rounded mean error score range of 4 to 28. Both "live" and "recorded" test administrations are represented at most of these various levels of achievement, indicating that sample differences and not administration method were largely responsible for fluctuations in achievement.

In general, those samples with lowest mean error scores also showed least variability of test performance as measured by standard deviations and ranges. These data are summarized in Table 15.

TABLE 15

Means, Standard Deviations (SD), and Ranges for Error Scores Achieved on USES Dictation Exercises No.4 and No.10 by the Ten Experimental Samples

Sample	N	Mean		SD		Range	
		Ex.4	Ex.10	Ex.4	Ex.10	Ex.4	Ex.10
1	25	24.080	28.320	15.515	18.371	6-68	4-75
2	35	9.257	9.086	9.034	8.016	0-39	1-33
3	29	20.276	20.000	14.790	13.807	4-57	2-48
4	33	5.212	6.091	4.689	6.012	1-21	0-23
5	18	8.167	5.667	8.234	4.000	1-28	0-15
6	34	4.559	6.000	6.386	8.967	0-28	0-36
7	11	5.818	4.273	4.933	3.910	0-15	0-13
8	31	12.258	13.387	10.746	11.878	1-47	1-40
9	14	18.071	17.857	18.737	19.654	1-74	1-68
10	12	17.250	25.333	18.749	20.536	1-66	4-72

Separating the ten samples into student and job applicant groups revealed the fact that performance on each exercise was similar for a given group. This grouping also showed that mean error scores for students were considerably lower than mean error scores for job applicants. Total error scores for the two exercises ranged from 0 to 74 for Exercise No. 4 and from 0 to 75 for Exercise No. 10. Table 16 presents these comparative data.

TABLE 16

Means, Standard Deviations (SD), and Ranges for Error Scores Achieved on USES Dictation Exercises No.4 and No.10 by Combined Experimental Samples

Exercise 4

Sample	N	Mean	SD	Range
BYU	216	11.259	12.223	0-68
PCS	26	17.692	18.747	1-74
Total	242	11.950	13.231	0-74

Exercise 10

Sample	N	Mean	SD	Range
BYU	216	11.921	13.192	0-75
PCS	26	21.308	20.409	1-72
Total	242	12.930	14.441	0-75

When the various BYU student samples were combined along the specific research dimensions of number of examiners, method of administration, and stenographic experience, performance differences were reduced. Table 17 shows means, standard deviations, and ranges for error scores achieved on Exercises 4 and 10 for these combined samples. Examination of this table reveals that a given student sample had similar mean scores and similar standard deviations on both dictation exercises.

TABLE 17

Means, Standard Deviations, (SD), and Ranges for Error Scores Achieved on USES Dictation Exercises No. 4 and No. 10 by Combined Samples of Brigham Young University Students

Sample	N	Means		SD		Range	
		Ex. 4	Ex. 10	Ex. 4	Ex.10	Ex.4	Ex.10
Live-One Examiner	46	8.435	7.93[illegible]	8.371	7.534	0-39	0-33
Live-Two Examiners	59	12.831	15.458	14.783	17.635	0-68	0-75
All Live	105	10.905	12.162	12.580	14.613	0-68	0-75
All Recorded	111	11.595	11.694	11.864	11.687	1-57	0-48
Live-Experience	50	8.320	9.680	8.599	12.791	0-31	0-58
Live-No Experience	55	13.255	14.418	14.944	15.756	0-68	0-75
Recorded-Experience	42	7.624	5.595	8.531	4.440	1-44	0-21
Recorded-No Experience	69	14.377	15.406	12.716	13.089	1-57	0-48

Differences between mean error scores achieved on the two exercises by each combined student sample were tested for significance. "t" ratios showed no significant difference in mean performance on the two exercises for any of these samples (see Table 18). On the other hand, there were some significant differences in mean performance when the various student samples were compared. These differences were limited as far as method of administration and number of examiners were concerned. On Exercise No. 10 the mean error score for the "live" group with two examiners was greater than that for the "live" group with one examiner at the .01 level of significance.

TABLE 18

Significance of Difference between Mean Error Scores Achieved by Each Combined Sample of Brigham Young University Students on USES Dictation Exercises No. 4 and No. 10

Sample	Critical Ratio	Significance Level
Live-One Examiner	-1.645	Not Sig.
Live-Two Examiners	.462	Not Sig.
All Live	-1.221	Not Sig.
All Recorded	- .148	Not Sig.
Live-Experience	-1.075	Not Sig.
Live-No Experience	- .734	Not Sig.
Recorded-Experience	1.541	Not Sig.
Recorded-No Experience	-1.143	Not Sig.

More differences existed when the factor of experience was applied. On Exercise No. 4 the mean error score for the "live" group without experience was significantly higher than the "live" group with experience. The mean error score on this same test for "recorded" group without experience was also significantly higher than the mean error score for the "recorded" group with experience. Significance levels for these comparisons were .05 and .01 respectively. On Exercise No. 10 two significant differences were found when comparing mean performance of experienced and nonexperienced groups. On the average, the recorded sample without experience made significantly more errors than the recorded group with experience. The significance level of this difference was .01. Also, the "live" group with experience had a significantly higher mean error score (.05 level of significance) than its recorded counterpart. These differences are shown in Table 19.

TABLE 19

Significance of Difference between Mean Error Scores Achieved by Combined Samples of Brigham Young University Students on Dictation Exercises No. 4 and No. 10

Sample Comparison	Critical Ratio	Significance Level
Live-One Examiner #4 vs Live-Two Examiners #4	-1.923	Not Sig.
Live-One Examiner #10 vs Live-Two Examiners #10	-2.950	.01
All Live #4 vs All Recorded #4	- .414	Not Sig.
All Live #10 vs All Recorded #10	.259	Not Sig.
Live-Experience #4 vs Live-No Experience #4	-2.096	.05
Live-Experience #10 vs Live-No Experience #10	-1.698	Not Sig.
Recorded-Experience #4 vs Recorded-No Experience #4	-3.642	.01
Recorded-Experience #10 vs Recorded-No Experience #10	-5.711	.01
Live-Experience #4 vs Recorded-Experience #4	.723	Not Sig.
Live-Experience #10 vs Recorded-Experience #10	2.112	.05
Live-No Experience #4 vs Recorded-No Experience #4	- .443	Not Sig.
Live-No Experience #10 vs Recorded-No Experience #10	- .374	Not Sig.

An analysis of variance for the three factors of tests, method of administration, and experience was made for the total BYU student sample. Table 20 summarizes the results of this analysis. These data indicate that experience is the only factor that exerted any significant influence on shorthand skill. The "F" ratio of 17.769 for this factor is significant at the .001 level.

TABLE 20

Summary of Analysis of Variance

Source of Variation	df	SS	MS	F
Between subjects	215			
A (Method)	1	69.47961	69.47961	.262
B (Experience)	1	4709.63432	4709.63432	17.769*
AB (Interaction A x B)	1	367.74323	367.74323	1.387
Denominator Term	212	56189.87650	265.04659	
Within subjects	215			
C (Tests)	1	29.93986	29.93986	.089
AC (Interaction A x C)	1	55.74268	55.74268	.166
BC (Interaction B x C)	1	33.39174	33.39174	.100
ABC (Interaction A x B x C)	1	45.56486	45.56486	.136
Denominator Term	211	70681.80690	334.98487	

*Significant at the .001 level

Some interesting differences in mean performance of student samples are evident from these statistical analyses. First of all, there is a definite trend for the "live" sample with one examiner to exhibit better average performance on each dictation exercise than the "live" sample with two examiners. The mean difference on Exercise 4 approaches significance at the .05 level,

and the mean difference on Exercise 10 is significant at the .01 level. Secondly, the method of administration--"live" or "recorded"--has little bearing on average performance. Last of all, the factor of experience exerts a greater influence on test performance than either the test form or the method of administration.

"t" ratios computed for the job applicant samples indicated there was no significant difference in mean performance on the two exercises for Sample 9. However, the mean number of errors on Exercise 10 was significantly greater than the mean number of errors on Exercise 4 at the .01 level for Sample 10. Cross comparisons between the two samples revealed no significant differences in mean performance on Exercise 4 and Exercise 10.

Test Performance and Personnel Factors

The relationships between test performance and age, education, shorthand training, and shorthand work experience were determined by computing Pearson product-moment correlations with error scores for each personnel factor. Table 21 shows these correlations for the ten experimental samples. Only a few of these correlation coefficients were significant. Age correlated significantly with Sample 6 error scores at the .05 level for Exercise 4 and at the .01 level for Exercise 10. Years of education correlated significantly with error scores for Sample 6 at the .01 level for Exercise 10, and with error scores for Sample 8 at the .05 level for Exercise 4.

This general lack of significant relationship between test performance and age and education is not too surprising when one takes into consideration the fact that these experimental samples were not radically different as far as means and standard deviations for age and education are concerned (see Tables 3 and 4). Furthermore, the amount of general education does not appear to be a critical factor after a certain minimum level has been reached.

TABLE 21

Pearson Product-Moment Correlations with USES Dictation Test Error Scores for Age, Education, Shorthand Training and Shorthand Work Experience for the Ten Experimental Samples

Sample	N	Age (years)		Education (years)		Training (no. of courses)		Experience (months)	
		Ex. 4	Ex. 10	Ex. 4	Ex. 10	Ex. 4	Ex. 10	Ex. 4	Ex. 10
1	25	.039	-.177	-.020	-.316	-.283	-.288	-.226	-.082
2	35	-.158	-.023	-.171	.137	.074	.021	-.148	-.175
3	29	-.116	-.139	-.029	-.034	.099	-.010	.097	-.142
4	33	-.135	.045	-.140	.134	.021	-.081	-.015	.017
5	18	.428	.211	.457	.283	.305	.309	.275	.122
6	34	.355*	.509**	.314	.521**	-.056	-.009	-.170	-.200
7	11	.208	-.448	.366	-.508	-.311	.453	-.224	.345
8	31	.253	.080	.418*	.344	-.129	-.185	-.281	-.415*
9	14	-.223	-.129	-.046	-.017	-.391	-.542*	-.235	-.142
10	12	-.134	-.114	-.119	-.074	-.744**	-.607*	-.124	-.160

*Significant at the .05 level
**Significant at the .01 level

Significant negative correlations between error scores and shorthand training and shorthand work experience were also few in number. The number of shorthand courses correlated significantly with three of the four error scores for job applicant Samples 9 and 10. On Exercise 4 the error scores of Sample 10 correlated with training at the .01 level of significance, and on Exercise 10 the error scores of both Samples 9 and 10 correlated with training at the .05 level of significance. Only one significant negative correlation was found between length of shorthand work experience and error scores. This occurred for Sample 8 error scores on Exercise 10 and was significant at the .05 level.

The above results indicate that there is no consistent relationship between test performance and personnel factors for the ten experimental samples. However, combining student Samples 1 through 8 according to method and experience gave a somewhat different picture. There were no significant correlations between age and test performance and only one significant negative correlation between education and test performance. The latter was at the .05 level for the Exercise 10 performance of the recorded group with no experience. Two negative correlations between work experience and test performance were significant. These were for Exercise 4 for the "all live" sample and for Exercise 10 for the "all recorded" sample. These negative correlations were significant at the .05 and .01 levels of significance respectively.

TABLE 22

Pearson Product-Moment Correlations with USES Dictation Test Error Scores for Age, Education, Shorthand Training and Shorthand Work Experience for Combined Samples of Brigham Young University Students

Sample	N	Age (years)		Education (years)		Training (no.of courses)		Experience (months)	
		Ex. 4	Ex. 10	Ex. 4	Ex.10	Ex. 4	Ex. 10	Ex. 4	Ex.10
One Examiner	46	-.122	-.083	-.111	-.009	.074	.172	-.132	-.099
Two Examiners	59	.244	.145	.157	.051	-.519**	-.493**	-.242	-.191
All Live	105	.133	.072	.021	-.025	-.323**	-.296**	-.199*	-.161
All Recorded	111	.083	.019	.154	.144	-.148	-.211*	-.170	-.249**
Live-Experience	50	.019	-.076	.064	-.066	-.295*	-.166	-.201	-.122
Live-No Experience	55	.170	.141	-.005	.001	-.306*	-.349**	----	----
Recorded-Experience	42	.111	-.078	.133	.034	-.126	-.242	.123	.199
Recorded-No Experience	69	.148	.133	.217	.252*	-.053	-.096	----	----

*Significant at the .05 level
**Significant at the .01 level

The greatest number of significant negative correlations was found for length of shorthand training. For this variable, correlations which were significant at the .01 level occurred between Exercise 4 and Exercise 10 error scores for the "live" group having two examiners and for the "all live" group. Negative correlations significant at the .05 level were noted for the "all recorded" group on Exercise 10 and for the "live-experience" and "live-no experience" groups on Exercise 4. Exercise 10 error scores for the "live" group without experience also correlated negatively with training at the .01 level of significance. These results which are shown in Table 22 indicate that, although the pattern is not consistent across all combined student samples, there is a definite trend for students with more shorthand training to make fewer transcription errors. This trend also agrees with correlation results of the two job applicant samples.

Reliability of Testing Methods

The major purpose for conducting this research was to determine whether there is a significant difference in the reliability of administering dictation tests "live" or by recording. Recorded administration appears to have the greater consistency since the timing and delivery of each dictation exercise are the same for each testing session and the role of the examiner is less critical. "Live" administration seems to have greater possibilities for inconsistency because each test has to be timed as it is being given, individual examiners may vary their delivery slightly from group to group and from day to day, and the test may be administered by different examiners all having their own individual test administration style. Another contributing factor to unreliability of "live" administration is the less frequent administration of proficiency tests required in small public employment offices.

The apparent reliability advantage of recorded dictation tests is offset by certain limitations. Special equipment must be installed in each test center to play the recordings, the equipment and installation must provide acceptable sound reproduction, and the personal touch of "live" administration is lost. In this study an attempt was made to overcome these deficiencies by using good sound equipment and retaining "live" administration of test directions.

Reliability of testing methods was determined by computing reliability coefficients using error scores on Dictation Exercises 4 and 10. Table 23 shows these reliability coefficients for the ten experimental samples.

TABLE 23

Reliability Coefficients for Live and Recorded Dictation Exercises

BYU Student Samples

LIVE				RECORDED			
One Examiner		Two Examiners		Individual Examiners			
Sample 2	Sample 7	Sample 1	Sample 6	Sample 3	Sample 4	Sample 5	Sample 8
.601	.050	.520	.619	.789	.773	.803	.778
.577		.728					
.711				.820			

PCS Job Applicant Samples

LIVE	RECORDED
Sample 9	Sample 10
.862	.944

It is interesting to note that the coefficients for the "recorded" samples are, as a group, larger than the coefficients for the "live" samples; but, raw coefficient values can be misleading when determining significant differences in reliability. Consequently, these differences were analyzed by transforming the reliability coefficients to z coefficients, computing the standard error of difference between these transformed coefficients and then checking the critical ratios between the differences of these coefficients and their standard errors for significance. Critical ratios of reliability differences for the ten samples are presented in Table 24.

TABLE 24

Critical Ratios of Differences between Reliability Coefficients for Mean Error Scores Achieved by the Ten Experimental Samples on USES Dictation Exercises No. 4 and No. 10

	First Element of Paired Comparisons							
Sample	1	2	3	4	5	6	7	9
2	-.397							
3	-1.692	-1.439						
4	-1.568	-1.299	.187					
5	-1.553	-1.310	.0925	- .253				
6	- .538	- .159	1.279	1.132	1.176			
7	1.284	1.619	2.523*	2.438*	2.398*	1.715		
8	-1.650	-1.391	.0734	- .114	.156	-1.227	-2.494*	
10	----	----	----	----	----	----	----	-1.0013

*Significant at the .05 level

These critical ratios indicate that reliability was not affected significantly by method of test administration. Sample 7 was significantly less reliable than Samples 3, 4, 5 and 8 at the .05 level. However, the low reliability of Sample 7 (.050) may be explained in part by the chance fluctuations of a small sample. Although the greatest error score for this sample was 15, the complete reversal of error scores (some positive and some negative changes) by six of the 11 sample members reduced the reliability coefficient almost to zero.

Because the "n's" for each of the ten experimental samples were small, reliability coefficients were also computed for the combined samples of Brigham Young University students. These reliability data are included in Table 23. This table shows that Sample 7 influenced the reliability of the "live" group with one examiner, resulting in a reliability coefficient of .577 compared with one of .728 for the "live" group with two examiners. However, the difference between these reliability statistics is not significant.

The same trend for greater stability of recorded dictation tests noted for individual samples persisted for the combined samples. The two reliability coefficients are .711 and .820 for the "all live" and "all recorded" groups respectively. The "t" ratio for the difference between these two groups is 1.95 which is close to the 1.96 value required for the .05 level of significance.

Table 25 shows critical ratios and significance levels for these reliability coefficient comparisons.

TABLE 25

Significance of Difference between Reliability Coefficients of Mean Error Scores Achieved on USES Dictation Exercises No. 4 and No. 10 by Combined Samples of Brigham Young University Students

Sample Comparison	Critical Ratio	Significance Level
Live-One Examiner vs Live-Two Examiners	-1.332	Not Sig.
All Live vs All Recorded	-1.954	Not Sig.

A final look at "live" and "recorded" test reliability for student samples was made by adding the factor of experience. Reliability coefficients of "live" and "recorded" dictation tests administered to students with shorthand work experience and to students without shorthand work experience appear in Table 26.

TABLE 26

Reliability Coefficients for Live and Recorded Dictation Exercises Administered to Brigham Young University Students Classified According to Experience

LIVE		RECORDED	
Experience	No Experience	Experience	No Experience
.717	.708	.746	.832

No significant difference was found between the reliability of the "live" group with experience and the "live" group without experience. For these two groups the coefficients were .717 and .708. Although the reliability coefficient for the recorded group without experience (.832) was greater than the

reliability coefficient for the recorded group with experience (.746), the difference between these coefficients was not significant. Comparisons between the "live" and "recorded" groups with experience and the "live" and "recorded" groups without experience also showed no significant differences in reliability. Table 27 summarizes the critical ratios of these reliability comparisons.

TABLE 27

Significance of Difference between Reliability Coefficients for Mean Error Scores Achieved on USES Dictation Exercises No. 4 and No. 10 by Combined Samples of Brigham Young University Students Classified According to Experience

Sample Comparison	Critical Ratio	Significance Level
Live-Experience vs Live-No Experience	.099	Not Sig.
Recorded-Experience vs Recorded-No Experience	-1.089	Not Sig.
Live-Experience vs Recorded-Experience	- .277	Not Sig.
Live-No Experience vs Recorded-No Experience	-1.618	Not Sig.

The above analysis shows that administration of dictation tests by the "recorded" method is fully as reliable as administration by the "live" method. In fact, reliability coefficients of "recorded" administrations are larger, although not significantly so, than those of "live" administrations.

SUMMARY AND CONCLUSIONS

This research was conducted to determine the comparative reliability of "live" and "recorded" dictation tests. The total experimental sample of 242 examinees was drawn from (1) University students and (2) local office job applicants. Each sample had only one male participant. The 216 students were enrolled in a second semester shorthand course at Brigham Young University, Provo, Utah. Most of the instruction in this course was given by female instructors, but recordings with a male voice were used regularly in the teaching process. The 26 job applicants were applying for stenographic positions through the Professional, Clerical, and Sales Office of the Utah Department of Employment Security, Salt Lake City, Utah. The average member of the student sample was 19.9 years of age, had 13.6 years of education, had taken 2.3 courses in shorthand, and had had 5.7 months of shorthand work experience. The average member of the job applicant sample was .7 years older, had .9 years less education, had completed .2 fewer shorthand courses, and had had more than four times as much on-job shorthand experience.

The experimental test battery included USES Dictation Exercise No. 4 and USES Dictation Exercise No. 10 taken from the Guide to the Use of Typing, Dictation and Spelling Tests, published by the Bureau of Employment Security in December, 1953. These exercises were dictated at 80 words per minute either "live" or by tape recording. The same standard tape recording with a male voice was used for "recorded" administrations. Standardized test directions were given "live" from the Guide. Supplementary instructions were given to clarify "recorded" administration. A brief introduction stating the purpose of testing and a practice exercise preceded the two test exercises.

Test performance was measured by number of transcription errors. Transcriptions were scored according to "Directions for Scoring Transcription" appearing in the Guide. Personnel information including age, education, amount of shorthand training, and length of shorthand work experience along with preferences for method of test administration and sex of test administrator were obtained through a special Information Sheet. Personnel data were collected from all members of the experimental sample, but preference data were lacking for one member of the job applicant sample.

Test data were collected on May 11 and 12, 1966 for the student sample. Students were tested in eight groups, ranging in size from 11 to 35 examinees. Four of these groups (105 students) were administered the test battery "live." The other four groups (111 students) were administered the test battery by tape recording. The "live" groups had either one or two examiners administering the tests to them, and the "recorded" groups had only one examiner administering the tests. Both male and female examiners were represented. Each exercise was fully transcribed after its dictation. Standard transcription time was reduced from 20 to 15 minutes for each exercise to accommodate the 50-minute class period.

Job applicants were tested from June 6 to 23, 1966. Two female examiners administered the tests to this sample in groups of one to four examinees. Method of administration was changed randomly from day to day. Fourteen members of the job applicant sample were given the test "live," and 12 were given the test by recording. Sometimes one examiner gave both tests; and sometimes both examiners assisted, each examiner administering one of the test exercises. Examinees in the job applicant sample were allowed the standard 20 minute period to transcribe each test exercise.

Test data were analyzed individually for the eight student subsamples and the two job applicant subsamples. They were also analyzed collectively for the combined student and combined job applicant samples. Testing preferences were analyzed not only for these samples but also for the combined total sample. This distinction in analytical comparison was made because of basic differences in personnel characteristics and testing conditions between the student and job applicant samples.

Quantitative analysis included computation of means, standard deviations, ranges, Pearson product-moment correlations, reliability coefficients, standard errors of differences between means and between reliability coefficients, analysis of variance, and t and F ratios. The comparing of so many differences between small samples through the use of "t" tests may be a questionable technique, but the findings from this procedure are supported by the results of analysis of variance.

The analysis of student and job applicant data supports the following conclusions:

1. Either "live" or "recorded" dictation test administration is acceptable to examinees since no clear-cut administration preference was expressed by either students or job applicants. Combined total sample preference percentages were 29 for "live," 30 for "recorded," and 41 for either method.
2. Most examinees prefer taking dictation from a male voice. For all ten samples 70% preferred a male voice, 5% preferred a female voice, and 25% had no preference.
3. Mean error scores on each dictation exercise were generally similar for a given subsample. They were also similar for a given group when student samples were combined along specific

research dimensions and when the ten subsamples were separated into student and job applicant samples.

4. Although students had less transcription time than job applicants, their mean error scores were considerably lower. Respective means were 11.3 and 17.7 for Exercise No. 4 and 11.9 and 21.3 for Exercise No. 10.
5. Differences between mean error scores achieved on the two exercises by each combined student samples were not significant. However, the mean error score on Exercise No. 10 was significantly higher than the mean error score on Exercise No. 4 for Applicant Sample 10.
6. There were some significant differences in mean performance for cross comparisons of combined student samples but not for job applicant samples. These differences were limited for method of administration but were more numerous when samples were classified according to experience.
7. Mean student performance of the "recorded" group with experience was significantly better on the two exercises than the mean performance of the "recorded" group without experience. Both of these differences are significant at the .01 level. Mean student performance of the "live" group with experience was significantly better on Exercise No. 4 than the mean performance of the "live" group without experience (.05 level of significance).
8. An analysis of student sample variance verifies the fact that experience rather than test form or method of administration exerted the greatest significant influence on shorthand skill.
9. As measured by Pearson product-moment correlations, there is no consistent relationship between test performance and the personnel factors of age, education, amount of shorthand training, and

and length of shorthand experience for the ten experimental samples. However, for combined "live" student samples there is a trend for those with more shorthand training to perform better on the dictation exercises.

10. Reliability coefficients are larger for individual subsamples administered the test battery by recording than for subsamples administered the test battery "live;" but the only significant differences in reliability occurred between Student Sample 7 (Live-One Examiner) and Student Samples 3, 4, 5, and 8 (Recorded). The low reliability of Sample 7 (.050) can be explained by the chance fluctuations in performance of a small sample (N=11). Consequently, these comparative statistics are of little import.
11. Larger reliability coefficients favoring the "all recorded" student sample over the "all live" student sample, the "live" student sample with two examiners over the "live" student sample with one examiner, and the "recorded" applicant sample over the "live" applicant sample were not statistically significant.
12. There were no significant differences in reliability among the "live" and "recorded" student samples when these were classified according to shorthand experience.
13. Although student examinees had the advantage of being familiar with recorded dictation tests, the administration of dictation tests by the "recorded" method appears to be fully as reliable as administration by the "live" method.

GPO 886-550